Available online at www.sciencedirect.com
ScienceDirect
J. Math. Anal. Appl. 343 (2008) 190–206
Journal of MATHEMATICAL ANALYSIS AND APPLICATIONS
www.elsevier.com/locate/jmaa

Improved entropy decay estimates for the heat equation

A. Arnold [a], J.A. Carrillo [b], C. Klapproth [c,*]

[a] *Institut für Analysis und Scientific Computing, E101, Technische Universität Wien, Wiedner Hauptstr. 8, A-1040 Wien, Austria*
[b] *Institució Catalana de Recerca i Estudis Avançats and Departament de Matemàtiques, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain*
[c] *Zuse Institute Berlin (ZIB), Takustr. 7, D-14195 Berlin, Germany*

Received 22 April 2007

Available online 26 January 2008

Submitted by P. Sacks

Abstract

Improved entropy decay estimates for the heat equation are obtained by selecting well-parametrized Gaussians. Either by mass centering or by fixing the second moments or the covariance matrix of the solution, relative entropy toward these Gaussians is shown to decay with better constants than classical estimates.

Keywords: Entropy method; Heat equation; Large time behavior

1. Introduction

Describing the asymptotic behavior of diffusion and homogeneous kinetic models has recently received a lot of attention in the partial differential equations community [2]. Several approaches have been established to determine decay estimates toward a distinguished profile for large times. In purely diffusive models, these asymptotic profiles are typically given by self-similar solutions usually coming from stationary solutions of equations in self-similar variables. The use of logarithmic entropies to study large time asymptotics is classical in kinetic theory [11] and it was brought up for diffusion equations in the seminal papers of G. Toscani [26,27].

Variations of the entropy–entropy dissipation method connected to the Bakry–Emery strategy [4,5] have been used to describe these rates for linear and non-linear diffusion equations [3,9,10]. Deep connections to optimal transport issues were discovered by F. Otto [24]. He obtained these decay estimates using a suitable interpretation of the diffusion equations as gradient flows/steepest descent of entropy or free-energy functionals with respect to a formal Riemannian structure inducing an optimal transport distance. These decay estimates in some cases were already known by classical techniques involving compactness, scalings and maximum principle arguments [6,29,30]. In all these works, the

* Corresponding author.
E-mail addresses: anton.arnold@tuwien.ac.at (A. Arnold), carrillo@mat.uab.es (J.A. Carrillo), klapproth@zib.de (C. Klapproth).

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2008.01.046

decay estimates have been obtained in different senses, mainly: entropy decay [3,9,10,24], optimal transport distance decay [24], and L^1-decay [29], and in [9,10,24] as a consequence of Csiszár–Kullback type inequalities [12,21,28].

Once the first asymptotic term has been pinpointed, the next step is to improve the decay rate either by taking into account other invariances of the equation or by identifying the next term in the large-time asymptotic expansion. In the case of the heat equation, expansion at all orders of the solutions for large times in L^1 were obtained in [14]. More precisely, as long as more and more moments of the initial data are bounded, a better approximation in terms of derivatives of the fundamental solution of the heat equation and the moments of the initial data can be given for large times. A similar result without identifying the asymptotic expansion at all orders was obtained in [16] by using Fourier-based distances.

Obtaining the next terms in the asymptotic expansion and identifying the corresponding improved decay estimates and rates are interesting and important open question in the non-linear diffusion case. This question has been addressed recently for the fast diffusion equation at the linearized level [13] and, finally, proved for the non-linear fast-diffusion equation in [18,19,22]. These results take advantage of the complete knowledge of the spectrum of the linearized operator: first they show that solutions will lie for large times in a neighborhood of an asymptotic profile and then, they try from the linearized improved decay rates to infer the result over the non-linear one. In particular, they show that mass-centering speeds up the convergence rate for different particular cases of the diffusion exponent. In the case of the porous medium equation, a formal expansion to all orders of the solutions in the one-dimensional case was done in [1]. Finally, the improvement of decay rates and decay estimates for the porous medium equation by either mass-centering of by fixing equal variance was discussed in [31]. Essentially, these results give decay improvements in L^1-spaces. An improvement on the optimal transport distance decay by mass centering has been reported in the one-dimensional case [8].

Here, we will show how the entropy decay estimates for the heat equation can be improved by mass-centering and by fixing the covariance matrix of the approximations. In Section 2, an improved decay estimate for the heat equation by fixing center of mass and variance is obtained, whereas Section 3 is devoted to generalize this idea in the case of fixing the whole covariance matrix of the approximated Gaussian. Let us finally mention that these improvements will be at the level of the *constants* in the decay estimates but not at the level of the *decay rates* for large times. Although the improvement of the decay rate is expected and true at the L^1 level [14], the present approach does not yield it for the relative entropy.

2. The heat equation and isotropic Gaussians

It is well known that solutions of the Cauchy problem for the heat equation with diffusion constant $k/2$,

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{k}{2}\Delta u, \quad x \in \mathbb{R}^n,\ t > 0, \\ u(x, t = 0) = u_0(x), \quad x \in \mathbb{R}^n, \end{cases} \tag{2.1}$$

behave asymptotically like a Gaussian with the same mass as the solution and a variance that is linearly increasing in t. This result can be easily recovered from the classical logarithmic Sobolev inequality (LSI) of Gross [3,17] in $\mathbb{R}^n$ with respect to the isotropic Gaussian measure

$$dM_\sigma = (2\pi\sigma)^{-\frac{n}{2}} e^{-\frac{|x|^2}{2\sigma}}\, dx \tag{2.2}$$

as in [26] guided by classical arguments from kinetic theory. Let us quickly review a simplified proof of the one given in [26]. Consider any two probability densities ρ_1, ρ_2 on $\mathbb{R}^n$, i.e. $\rho_1, \rho_2 \in L^1_+(\mathbb{R}^n)$ with

$$\int_{\mathbb{R}^n} \rho_1\, dx = \int_{\mathbb{R}^n} \rho_2\, dx = 1.$$

We define the *relative logarithmic entropy* of ρ_1 w.r.t. ρ_2 as

$$e_1(\rho_1|\rho_2) := \int_{\mathbb{R}^n} \frac{\rho_1}{\rho_2} \ln \frac{\rho_1}{\rho_2} \rho_2\, dx \geqslant 0,$$

and $\int \rho_1 \ln \rho_1 \, dx$ is the *logarithmic entropy* of the probability density ρ_1. The LSI w.r.t. the Gaussian measure M_σ reads [17]

$$\int_{\mathbb{R}^n} g^2 \ln g^2 \, dM_\sigma \leqslant 2\sigma \int_{\mathbb{R}^n} |\nabla g|^2 \, dM_\sigma, \tag{2.3}$$

for all $\sigma > 0$ and $g \in L^2(\mathbb{R}^n, dM_\sigma)$ with $\int g^2 \, dM_\sigma = 1$. By setting $g^2 = \frac{\rho}{M_\sigma}$ it is equivalent to

$$e_1(\rho|M_\sigma) \leqslant \frac{\sigma}{2} I(\rho|M_\sigma), \tag{2.4}$$

for all $\rho \in L^1_+(\mathbb{R}^n)$ with $\int \rho \, dx = 1$. Here,

$$I(\rho_1|\rho_2) := \int_{\mathbb{R}^n} \frac{\rho_2}{\rho_1} \left| \nabla \left(\frac{\rho_1}{\rho_2} \right) \right|^2 \rho_2 \, dx = \int_{\mathbb{R}^n} \left| \nabla \ln \left(\frac{\rho_1}{\rho_2} \right) \right|^2 \rho_1 \, dx = 4 \int_{\mathbb{R}^n} \left| \nabla \sqrt{\frac{\rho_1}{\rho_2}} \right|^2 \rho_2 \, dx \geqslant 0 \tag{2.5}$$

denotes the *relative Fisher information* of ρ_1 w.r.t. ρ_2 [15]. There is equality in (2.3) if and only if

$$g(x) = g_y(x) := \exp\left(\frac{x \cdot y}{\sqrt{\sigma}} - |y|^2 \right) \tag{2.6}$$

for an arbitrary $y \in \mathbb{R}^n$, as established by Carlen in [7].

Theorem 1 *(Standard decay estimate). (See [26].) Let the initial value for the heat equation $u_0 \in C(\mathbb{R}^n) \cap W^{1,2}_{\mathrm{loc}}(\mathbb{R}^n) \cap L^1_+(\mathbb{R}^n)$ be a probability density on $\mathbb{R}^n$ with finite second moment and entropy, i.e. $u_0(x) \geqslant 0$, $\int u_0 \, dx = 1$, $\int |x|^2 u_0 \, dx < \infty$, and $\int u_0 |\ln u_0| \, dx < \infty$. Then, the relative logarithmic entropy of the solution u to* (2.1) *w.r.t. $u_\infty(x,t) := M_{E+kt}(x - \tilde{x}_0)$ with an arbitrary $\tilde{x}_0 \in \mathbb{R}^n$ and an arbitrary $E > 0$ satisfies the decay estimate*

$$e_1\big(u(t)\big|u_\infty(t)\big) \leqslant \frac{E}{E+kt} e_1\big(u_0\big|u_\infty(0)\big), \quad \forall t \geqslant 0. \tag{2.7}$$

Proof. Since u and u_∞ are solutions of the heat equation, they are smooth, positive, and rapidly decaying functions for all $t > 0$. Thus, we find for all $t > 0$,

$$\begin{aligned}
\frac{d}{dt} e_1\big(u(t)\big|u_\infty(t)\big) &= \int_{\mathbb{R}^n} \frac{\partial u}{\partial t} \left[\ln\left(\frac{u}{u_\infty} \right) + 1 \right] dx - \int_{\mathbb{R}^n} \frac{\partial u_\infty}{\partial t} \frac{u}{u_\infty} \, dx \\
&= \frac{k}{2} \int_{\mathbb{R}^n} \Delta u \left[\ln\left(\frac{u}{u_\infty} \right) + 1 \right] dx - \frac{k}{2} \int_{\mathbb{R}^n} \Delta u_\infty \frac{u}{u_\infty} \, dx \\
&= -\frac{k}{2} \int_{\mathbb{R}^n} \frac{u_\infty}{u} \nabla u \cdot \nabla \left(\frac{u}{u_\infty} \right) dx + \frac{k}{2} \int_{\mathbb{R}^n} \nabla u_\infty \cdot \nabla \left(\frac{u}{u_\infty} \right) dx \\
&= -\frac{k}{2} I\big(u(t)\big|u_\infty(t)\big).
\end{aligned} \tag{2.8}$$

Hence, $e_1(u(,t)|u_\infty(t))$ is non-increasing in time. Conservation of mass for the heat equation shows that u is a probability density for $t > 0$. And since $u_\infty(t)$ is a Gaussian with second moment $n(E + kt)$, we infer from the LSI (2.3) that

$$e_1\big(u(t)\big|u_\infty(t)\big) \leqslant \frac{E+kt}{2} I\big(u(t)\big|u_\infty(t)\big).$$

Applying this bound to the right-hand side of (2.8) yields with Gronwall's lemma

$$e_1\big(u(t)\big|u_\infty(t)\big) \leqslant \frac{E}{E+kt} e_1\big(u_0\big|u_\infty(0)\big), \quad \forall t \geqslant 0.$$

Since

$$e_1\big(u_0\big|u_\infty(0)\big)=\int\limits_{\mathbb{R}^n}u_0(x)\ln u_0(x)\,dx+\frac{n}{2}\ln\big(2\pi(E+kt)\big)+\frac{1}{2(E+kt)}\int\limits_{\mathbb{R}^n}|x-\tilde{x}_0|^2u_0(x)\,dx<\infty,$$

we conclude that u converges in logarithmic entropy to u_∞ as $t\to\infty$. □

Remark 2 *(Sharpness).* Due to the translational invariance of the heat equation and the equality cases (2.6) of the LSI, the decay estimate (2.7) is sharp in the following sense: choosing $u_0(x)=M_E(x-\tilde{x}_0')$ for an arbitrary $\tilde{x}_0'\in\mathbb{R}^n$, the solution of the Cauchy problem for the heat equation is $u(x,t)=M_{E+kt}(x-\tilde{x}_0')$ and

$$e_1\big(u(\cdot,t)\big|M_{E+kt}(\cdot-\tilde{x}_0)\big)=\frac{|\tilde{x}_0-\tilde{x}_0'|^2}{2(E+kt)},\quad\forall t\geqslant 0,\tag{2.9}$$

leading to non-trivial equality in (2.7).

On the way to improving the decay rate in relative entropy for the solution of (2.1) we shall compare $u(t)$ to a better fitted Gaussian $M_{\sigma(t)}$—rather than to M_{E+kt}. Solutions of the heat equation obviously conserve the center of mass:

$$\int\limits_{\mathbb{R}^n}xu(x,t)\,dx=\int\limits_{\mathbb{R}^n}xu_0(x)\,dx:=x_0,\quad\forall t>0,\tag{2.10}$$

and linearly increase the second moment:

$$\int\limits_{\mathbb{R}^n}|x-x_0|^2u(x,t)\,dx=\int\limits_{\mathbb{R}^n}|x-x_0|^2u_0(x)\,dx+nkt:=\alpha+nkt,\quad\forall t>0.\tag{2.11}$$

Our first observation is that

$$e_1\big(u(\cdot,t)\big|M_{\frac{\alpha}{n}+kt}(\cdot-x_0)\big)=\min_{\substack{E>0\\ \tilde{x}_0\in\mathbb{R}^n}}e_1\big(u(\cdot,t)\big|M_{E+kt}(\cdot-\tilde{x}_0)\big),\tag{2.12}$$

as it can easily be checked just by working on the explicit expression of the relative entropy $e_1(u(\cdot,t)|M_{E+kt}(\cdot-\tilde{x}_0))$. In other words, the optimal Gaussian (in the sense of minimizing the relative entropy) for a given solution u at a fixed time $t\geqslant 0$ is given by the Gaussian with the same center of mass and variance as u, i.e. $M_{\frac{\alpha}{n}+kt}(x-x_0)$.

In the following, we want to discuss if the decay estimate (2.7) for the relative entropy of u w.r.t. to the optimal Gaussian $M_{\frac{\alpha}{n}+kt}(x-x_0)$ given by

$$e_1\big(u(\cdot,t)\big|M_{\frac{\alpha}{n}+kt}(\cdot-x_0)\big)\leqslant\frac{\frac{\alpha}{n}}{\frac{\alpha}{n}+kt}e_1\big(u_0\big|M_{\frac{\alpha}{n}}(\cdot-x_0)\big),\quad t\geqslant 0,\tag{2.13}$$

is sharp as well. Clearly, we have equality for

$$u_0(x)=M_{\frac{\alpha}{n}}(x-x_0),\tag{2.14}$$

but then both sides of (2.13) are zero. For other cases, equality in (2.13) for all $t\geqslant 0$ would imply equality at $t=0$ of both t-derivatives, i.e. $-\frac{k}{2}I=-\frac{n}{\alpha}ke_1$. But this is the LSI (2.4), which becomes an equality only for the shifted Gaussians (2.6). But then, the equality of moments in the optimized Gaussian (cf. (2.12)) leaves (2.14) as the only case. Hence, there exists no initial function u_0 satisfying the conditions of (2.12) such that there is non-zero equality in (2.13) and this decay estimate is not sharp anymore.

Now, we come back to Theorem 1 and the estimate

$$e_1\big(u(\cdot,t)\big|M_{E+kt}(\cdot-\tilde{x}_0)\big)\leqslant\frac{E}{E+kt}e_1\big(u_0\big|M_E(\cdot-\tilde{x}_0)\big)\tag{2.15}$$

for an arbitrary $\tilde{x}_0\in\mathbb{R}^n$ and an arbitrary $E>0$. We observe that for small values $E>0$ the ratio $E/(E+kt)$ decays faster to zero as t goes to infinity than for large E's. This leads us to the conjecture that it is possible to find a sharper estimate for the logarithmic entropy of the solution u w.r.t. the Gaussian $M_{\frac{\alpha}{n}+kt}(\cdot-x_0)$ than (2.13) by determining

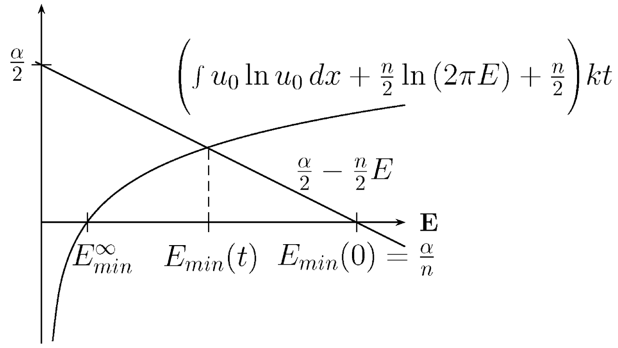


Fig. 1. $E_{\min}(t)$ is the intersection point of the functions on the left- and right-hand sides of (2.18) for all $t \geqslant 0$. $\lim_{t\to\infty} E_{\min}(t) = E^{\infty}_{\min}$ and $E_{\min}(0) = \frac{\alpha}{n}$ are the unique roots of the left- and right-hand sides of (2.18).

a function $E(t)$ with $0 < E(t) < \frac{\alpha}{n}$ for $t > 0$, which should be used instead of a constant E on the right-hand side of (2.15).

The idea for deriving such an optimized decay estimate is to minimize the right-hand side of inequality (2.15) w.r.t. $E > 0$ and $\tilde{x}_0 \in \mathbb{R}^n$. As before, we find for all fixed times $t \geqslant 0$ that

$$e_1\big(u(\cdot,t)\big|M_{\frac{\alpha}{n}+kt}(\cdot - x_0)\big) = \min_{\substack{E>0\\ \tilde{x}_0\in\mathbb{R}^n}} e_1\big(u(\cdot,t)\big|M_{E+kt}(\cdot - \tilde{x}_0)\big) \leqslant \inf_{\substack{E>0\\ \tilde{x}_0\in\mathbb{R}^n}} \frac{E}{E+kt} e_1\big(u_0\big|M_E(\cdot - \tilde{x}_0)\big)$$
$$= \inf_{E>0} \frac{E}{E+kt} e_1\big(u_0\big|M_E(\cdot - x_0)\big). \tag{2.16}$$

In the case $u_0(x) = M_{\frac{\alpha}{n}}(x - x_0)$, we have $e_1(u(\cdot,t)|M_{\frac{\alpha}{n}+kt}(\cdot - x_0)) = 0$ for all times $t \geqslant 0$ and it holds equality in the estimate (2.7). Hence, we obtain the minimum of the right-hand side of (2.16) and therefore the best estimate for the choice $E = \frac{\alpha}{n}$. In general we have to determine the time-dependent second moment $E(t) > 0$ by minimizing for each fixed $t > 0$ the function $f : \mathbb{R}^+ \times \mathbb{R}^+_0 \to \mathbb{R}^+_0$ defined as

$$f(E,t) := \frac{E}{E+kt} e_1\big(u_0\big|M_E(\cdot - x_0)\big) = \frac{E}{E+kt}\left(\int_{\mathbb{R}^n} u_0(x)\ln u_0(x)\, dx + \frac{n}{2}\ln(2\pi E) + \frac{\alpha}{2E}\right) \tag{2.17}$$

which leads us to the following result:

Lemma 3 *(Computation of $E_{\min}(t)$). (See Fig. 1.) Let $u_0(x) \neq M_{\frac{\alpha}{n}}(x - x_0)$ on a set of positive measure. Then the function $f(E,t)$ defined by (2.17) has w.r.t. $E > 0$ for all fixed $t \geqslant 0$ a unique minimum $E_{\min}(t)$ with the following properties*:

(a) *$E_{\min}(t)$ satisfies*

$$\left(\int_{\mathbb{R}^n} u_0(x)\ln u_0(x)\, dx + \frac{n}{2}\ln\big(2\pi E_{\min}(t)\big) + \frac{n}{2}\right)kt = \frac{\alpha}{2} - \frac{n}{2}E_{\min}(t), \quad \forall t \geqslant 0. \tag{2.18}$$

(b) $E_{\min}(0) = \frac{\alpha}{n}$.

(c)

$$E^{\infty}_{\min} := \lim_{t\to\infty} E_{\min}(t) = \frac{\alpha}{n}\exp\left(-\frac{2}{n}e_1\big(u_0\big|M_{\frac{\alpha}{n}}(\cdot - x_0)\big)\right) = \frac{1}{2\pi}\exp\left(-\frac{2}{n}\int_{\mathbb{R}^n} u_0(x)\ln u_0(x)\, dx - 1\right).$$

(d) *$E_{\min}(t)$ is strictly monotonic decreasing w.r.t. $t \geqslant 0$. In particular,*

$$0 < E_{\min}(t) < \frac{\alpha}{n}, \quad \forall t > 0.$$

Proof. $f(E,t)$ is for all times $t \geqslant 0$ differentiable w.r.t. E with

$$\begin{aligned}\frac{\partial}{\partial E} f(E,t) &= \frac{kt}{(E+kt)^2} e_1\big(u_0 \big| M_E(\cdot - x_0)\big) + \frac{E}{E+kt}\left(\frac{n}{2E} - \frac{\alpha}{2E^2}\right) \\ &= \frac{kt}{(E+kt)^2}\left(\int_{\mathbb{R}^n} u_0 \ln u_0 \, dx + \frac{n}{2}\ln(2\pi E)\right) + \frac{n}{2(E+kt)} - \frac{\alpha}{2(E+kt)^2} = \frac{\tilde f(E,t)}{(E+kt)^2} \end{aligned} \tag{2.19}$$

where

$$\tilde f(E,t) = \left(\int_{\mathbb{R}^n} u_0 \ln u_0 \, dx + \frac{n}{2}\ln(2\pi E)\right)kt + \frac{n}{2}(E+kt) - \frac{\alpha}{2}$$

is monotonically increasing w.r.t. $E>0$. We observe that $\tilde f(E,t)$ converges to $-\infty$ as $E \to 0+$ and evaluating the function at the point $E = \frac{\alpha}{n}$ leads to

$$\tilde f\left(E = \frac{\alpha}{n}, t\right) = e_1\big(u_0 \big| M_{\frac{\alpha}{n}}(\cdot - x_0)\big) \cdot kt > 0, \quad \forall t > 0.$$

We conclude by the continuity of $\tilde f(E,t)$ w.r.t. $E>0$ and the intermediate value theorem that $\tilde f(E,t)$ is zero at one point $E_{\min}(t) \in (0, \frac{\alpha}{n})$ and thus, $\tilde f(E,t)$ and $\partial_E f(E,t)$ have a unique zero for $t>0$. The fact that $\tilde f(E,t)$ and $\partial_E f(E,t)$ have the same sign yields the uniqueness of a minimum $E_{\min}(t)$ in $(0,\frac{\alpha}{n})$ of $f(E,t)$ for $t>0$.

(a) By setting (2.19) to zero and rewriting the equation we find for all $t \geqslant 0$ a condition for $E_{\min}(t)$ such that of $\partial_E f(E,t)$ becomes zero

$$\left(\int_{\mathbb{R}^n} u_0 \ln u_0 \, dx + \frac{n}{2}\ln\big(2\pi E_{\min}(t)\big) + \frac{n}{2}\right)kt = \frac{\alpha}{2} - \frac{n}{2}E_{\min}(t). \tag{2.20}$$

(b) Evaluating (2.20) at $t=0$ yields $E_{\min}(0) = \frac{\alpha}{n}$.

(c) Since $E_{\min}(t)$ is bounded we find by (2.20) that $E_{\min}^\infty := \lim_{t\to\infty} E_{\min}(t)$ solves the equation

$$\int_{\mathbb{R}^n} u_0 \ln u_0 \, dx + \frac{n}{2}\ln\big(2\pi E_{\min}^\infty\big) + \frac{n}{2} = 0,$$

which is equivalent to

$$e_1\big(u_0 \big| M_{\frac{\alpha}{n}}(\cdot - x_0)\big) - \frac{n}{2}\ln\left(2\pi\frac{\alpha}{n}\right) + \frac{n}{2}\ln\big(2\pi E_{\min}^\infty\big) = 0. \tag{2.21}$$

(d) Differentiating the expression (2.20) w.r.t. $t \geqslant 0$ gives

$$E_{\min}'(t) = -\frac{E_{\min}(t)}{E_{\min}(t)+kt}\frac{2}{nt}\left(\frac{\alpha}{2} - \frac{n}{2}E_{\min}(t)\right), \quad \forall t>0.$$

Since $E_{\min}(t) \in (0,\frac{\alpha}{n})$ for $t>0$, we have $E_{\min}'(t) < 0$ for $t>0$ and

$$E_{\min}'(0) = -\frac{2k}{n} e_1\big(u_0 \big| M_{\frac{\alpha}{n}}(\cdot - x_0)\big) < 0. \tag{2.22}$$

Thus, $E_{\min}(t)$ is strictly monotonic decreasing w.r.t. $t \geqslant 0$.

This concludes the proof. □

In the case $u_0(x) = M_{\frac{\alpha}{n}}(x - x_0)$ a.e., we define $E_{\min}(t) := \frac{\alpha}{n}$. Using Lemma 3 together with (2.16) we can now improve the decay estimate for the relative entropy in Theorem 1.

Theorem 4 *(Improved decay estimate). Let the initial value $u_0 \in C(\mathbb{R}^n) \cap W^{1,2}_{\mathrm{loc}}(\mathbb{R}^n) \cap L^1_+(\mathbb{R}^n)$ be a probability density on $\mathbb{R}^n$ with finite second moment and finite absolute entropy, i.e. $u_0 \geqslant 0$, $\int u_0 \, dx = 1$, $\int |x|^2 u_0 \, dx < \infty$, and*

$\int u_0|\ln u_0|\,dx < \infty$. *Then the solution u of the initial value problem* (*IVP*) (2.1) *satisfies*

$$e_1\big(u(\cdot,t)\big|M_{\frac{\alpha}{n}+kt}(\cdot-x_0)\big) \leqslant \frac{E_{\min}(t)}{E_{\min}(t)+kt} e_1\big(u_0\big|M_{E_{\min}(t)}(\cdot-x_0)\big) = \frac{1}{kt}\left(\frac{\alpha}{2}-\frac{n}{2}E_{\min}(t)\right), \quad \forall t>0. \tag{2.23}$$

Proof. By (2.16) we obtain the optimized estimate

$$e_1\big(u(\cdot,t)\big|M_{\frac{\alpha}{n}+kt}(\cdot-x_0)\big) \leqslant f\big(E_{\min}(t),t\big).$$

Inserting the conditional equation (2.20) for $E_{\min}(t)$ in the definition of $f(E,t)$ gives

$$\begin{aligned} f\big(E_{\min}(t),t\big) &= \frac{E_{\min}(t)}{E_{\min}(t)+kt} e_1\big(u_0\big|M_{E_{\min}(t)}(\cdot-x_0)\big) \\ &= \frac{E_{\min}(t)}{E_{\min}(t)+kt}\left[\frac{1}{kt}\left(\frac{\alpha}{2}-\frac{n}{2}E_{\min}(t)\right)-\frac{n}{2}+\frac{\alpha}{2E_{\min}(t)}\right] \\ &= \frac{E_{\min}(t)}{E_{\min}(t)+kt}\left(\frac{1}{kt}+\frac{1}{E_{\min}(t)}\right)\left(\frac{\alpha}{2}-\frac{n}{2}E_{\min}(t)\right) \\ &= \frac{1}{kt}\left(\frac{\alpha}{2}-\frac{n}{2}E_{\min}(t)\right), \end{aligned}$$

concluding the proof. □

We end this section by analyzing the sharpness of the optimized decay estimate (2.23) from Theorem 4. For some fixed initial value u_0 let us assume that there holds equality in (2.23) on some (possibly small) time interval $[0,T]$, i.e.

$$e_1\big(u(\cdot,t)\big|M_{\frac{\alpha}{n}+kt}(\cdot-x_0)\big) = \frac{E_{\min}(t)}{E_{\min}(t)+kt}\left(\int_{\mathbb{R}^n} u_0\ln u_0\,dx + \frac{n}{2}\ln\big(2\pi E_{\min}(t)\big) + \frac{\alpha}{2E_{\min}(t)}\right).$$

Differentiating this equality w.r.t. the time $t \geqslant 0$ yields

$$\begin{aligned} \frac{d}{dt}e_1\big(u(\cdot,t)\big|M_{\frac{\alpha}{n}+kt}(\cdot-x_0)\big) &= \frac{E'_{\min}(t)kt - E_{\min}(t)k}{(E_{\min}(t)+kt)^2} e_1\big(u_0\big|M_{E_{\min}(t)}(\cdot-x_0)\big) \\ &\quad + \frac{E'_{\min}(t)}{E_{\min}(t)+kt}\left(\frac{n}{2}-\frac{\alpha}{2E_{\min}(t)}\right). \end{aligned} \tag{2.24}$$

From (2.8) we know that for all times $t \geqslant 0$,

$$\frac{d}{dt}e_1\big(u(\cdot,t)\big|M_{\frac{\alpha}{n}+kt}(\cdot-x_0)\big) = -\frac{k}{2} I\big(u(\cdot,t)\big|M_{\frac{\alpha}{n}+kt}(\cdot-x_0)\big).$$

Evaluating (2.24) at time $t=0$ gives with $E_{\min}(0)=\frac{\alpha}{n}$,

$$\frac{\alpha}{2n} I\big(u_0\big|M_{\frac{\alpha}{n}}(\cdot-x_0)\big) = e_1\big(u_0\big|M_{\frac{\alpha}{n}}(\cdot-x_0)\big).$$

This LSI becomes an equality only in the case (2.6), where

$$g^2(x) := \frac{u_0(x)}{M_{\frac{\alpha}{n}}(x-x_0)}.$$

Finally, we find with $t=0$ and $E=\frac{\alpha}{n}$ for the normalized u_0 the condition

$$u_0(x) = M_{\frac{\alpha}{n}}\left(x - \left(x_0 + 2\sqrt{\frac{\alpha}{n}}y\right)\right)$$

where $y \in \mathbb{R}^n$ is arbitrary. Since the first moment of u_0 is assumed to be equal to x_0, we conclude that

$$u_0(x) = M_{\frac{\alpha}{n}}(x-x_0).$$

Hence, we have equality in (2.23) only in the case that both sides of the inequality are equal to zero. So, the optimized decay estimate is not sharp in the sense above.

Next we compare our improved decay estimate (2.23) to the classical estimate (2.13) by Toscani [26]. Let us first discuss their large-time behaviors. Actually, (2.18) is equivalent to

$$\begin{aligned} E_{\min}(t) &= \frac{1}{2\pi}\exp\biggl[-\frac{2}{n}\int_{\mathbb{R}^n} u_0\ln u_0\,dx - 1 + \frac{2}{nkt}\biggl(\frac{\alpha}{2}-\frac{n}{2}E_{\min}(t)\biggr)\biggr] \\ &= E_{\min}^{\infty}\exp\biggl[\frac{2}{nkt}\biggl(\frac{\alpha}{2}-\frac{n}{2}E_{\min}(t)\biggr)\biggr], \quad \forall t>0, \end{aligned}$$

that can be expanded for $t \gg 1$,

$$E_{\min}(t) = E_{\min}^{\infty}\biggl[1+\frac{2}{nkt}\biggl(\frac{\alpha}{2}-\frac{n}{2}E_{\min}(t)\biggr)+\mathcal{O}\bigl(t^{-2}\bigr)\biggr].$$

Thus, the quantity $E_{\min}(t) - E_{\min}^{\infty}$ is for large times t proportional to $(E_{\min}^{\infty}+kt)^{-1}$ and we obtain for the improved decay estimate the approximation

$$\begin{aligned} e_1\bigl(u(\cdot,t)\big|M_{\frac{\alpha}{n}+kt}(\cdot-x_0)\bigr) &\leqslant \frac{1}{kt}\biggl[\frac{\alpha}{2}-\frac{n}{2}E_{\min}^{\infty}-\biggl(\frac{\alpha}{2}-\frac{n}{2}E_{\min}^{\infty}\biggr)\frac{E_{\min}^{\infty}}{E_{\min}^{\infty}+kt}+\mathcal{O}\bigl(t^{-2}\bigr)\biggr] \\ &= \frac{\frac{\alpha}{2}-\frac{n}{2}E_{\min}^{\infty}}{E_{\min}^{\infty}+kt}+\mathcal{O}\bigl(t^{-3}\bigr), \quad \forall t \gg 1. \end{aligned} \tag{2.25}$$

The large time behavior of our new decay estimate is similar to the original decay estimate: The right-hand side of (2.23) is proportional to $(E_{\min}^{\infty}+kt)^{-1}$ for large times t while the estimate (2.13) is proportional to $(\frac{\alpha}{n}+kt)^{-1}$.

Therefore, our new decay estimate does not improve the decay rate at ∞ that one may expect from centering and normalizing the Gaussian approximation (at least in L^1 and in weighted L^2-spaces). However, it does improve at the level of constants of decay, i.e., at the level of the ratio of the improved decay estimate (2.23) to the original estimate (2.13) w.r.t. to the time $t>0$. We define the function $R:\mathbb{R}^+\to\mathbb{R}^+$ describing this ratio by

$$R(t) := \frac{\frac{1}{kt}(\frac{\alpha}{2}-\frac{n}{2}E_{\min}(t))}{\frac{\frac{\alpha}{n}}{\frac{\alpha}{n}+kt}e_1(u_0|M_{\frac{\alpha}{n}}(\cdot-x_0))}.$$

Using the approximation (2.25) of $E_{\min}(t)$ for large times t gives

$$R(t) = \frac{\frac{\frac{\alpha}{2}-\frac{n}{2}E_{\min}^{\infty}}{E_{\min}^{\infty}+kt}}{\frac{\frac{\alpha}{n}}{\frac{\alpha}{n}+kt}e_1(u_0|M_{\frac{\alpha}{n}}(\cdot-x_0))}+\mathcal{O}\bigl(t^{-2}\bigr) = \frac{\frac{\alpha}{2}-\frac{n}{2}E_{\min}^{\infty}}{\frac{\alpha}{n}e_1(u_0|M_{\frac{\alpha}{n}}(\cdot-x_0))}\cdot\frac{\frac{\alpha}{n}+kt}{E_{\min}^{\infty}+kt}+\mathcal{O}\bigl(t^{-2}\bigr), \quad \forall t\gg 1. \tag{2.26}$$

We find that the ratio of the estimates is decreasing w.r.t. large times $t \gg 1$. Since the estimates coincide initially, the improvement of the decay rate becomes better for large times t and in the limit $t\to\infty$ the ratio converges with rate $(E_{\min}^{\infty}+kt)^{-1}$ to

$$R(\infty) = \frac{\frac{\alpha}{2}-\frac{n}{2}E_{\min}^{\infty}}{\frac{\alpha}{n}e_1(u_0|M_{\frac{\alpha}{n}}(\cdot-x_0))}. \tag{2.27}$$

Since $E_{\min}^{\infty} = \frac{\alpha}{n}\exp[-\frac{2}{n}e_1(u_0|M_{\frac{\alpha}{n}}(\cdot-x_0))]$ we find that the ratio $R(t)$ converges to

$$\frac{1-\exp[-\frac{2}{n}e_1(u_0|M_{\frac{\alpha}{n}}(\cdot-x_0))]}{\frac{2}{n}e_1(u_0|M_{\frac{\alpha}{n}})} \tag{2.28}$$

as the time t goes to infinity. This limit is monotonically decreasing for increasing logarithmic entropies $e_1(u_0|M_{\frac{\alpha}{n}}(\cdot-x_0))>0$. We finally point out that the function $E_{\min}(t)$ as defined by (2.18) only depends on the variance of u_0 and on its relative entropy w.r.t. the Gaussian $M_{\frac{\alpha}{n}}(\cdot-x_0)$. Fig. 2 shows the ratio function $R(t)$ for different times as a function of $\alpha := \int|x-x_0|^2u_0(x)\,dx$ and the relative entropy $e_1(u_0|M_{\frac{\alpha}{n}}(\cdot-x_0))$.

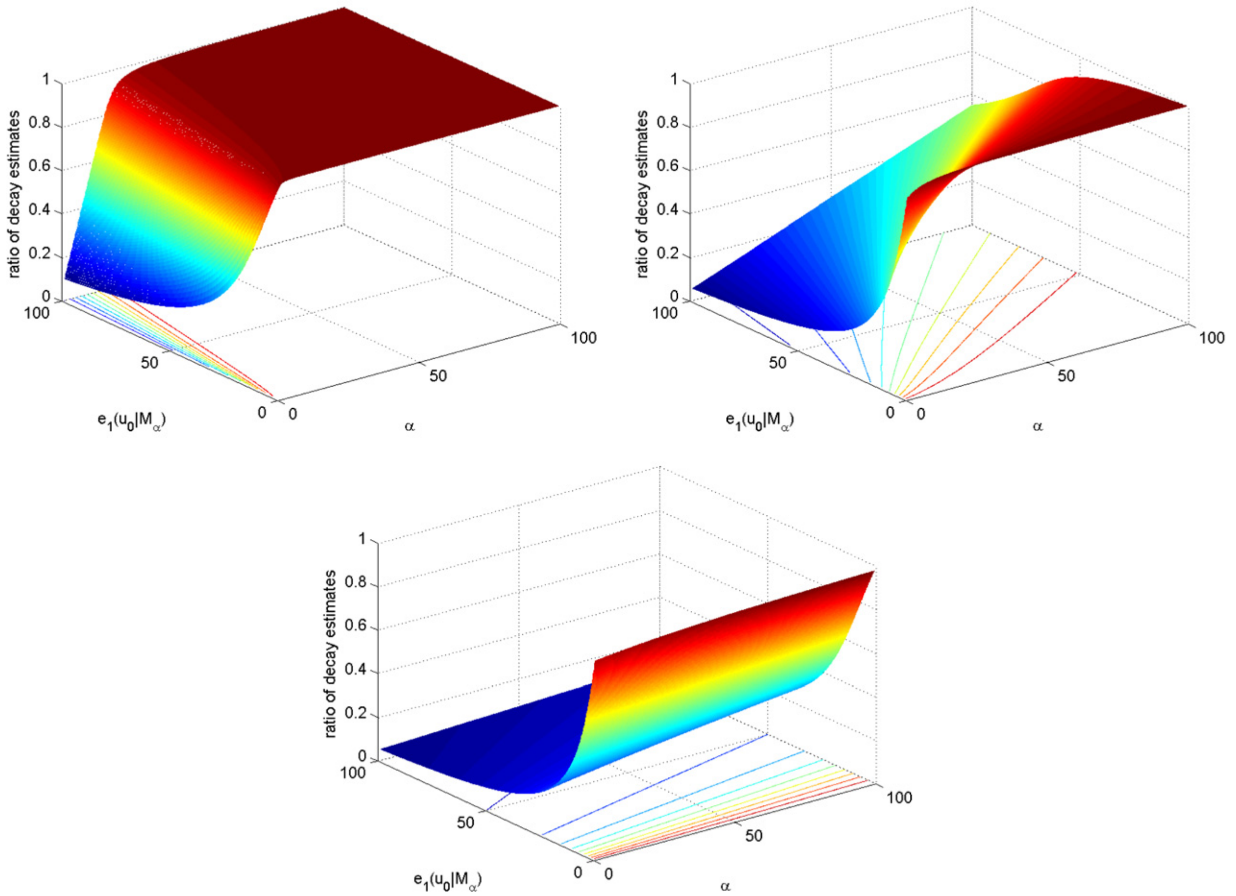


Fig. 2. $R(0.1)$, $R(1)$, and $R(10)$ for $n = 1$ and $k = 1$.

Example 5. We consider the initial function u_0 on $\mathbb{R}$ defined by

$$u_0(x) := \begin{cases} \frac{1}{2(x_2 - x_1)}, & x_1 \leqslant |x| \leqslant x_2, \\ 0, & \text{otherwise}, \end{cases} \tag{2.29}$$

where $x_2 > x_1 > 0$. It is an even probability density with second moment $\alpha = \frac{1}{3}(x_1^2 + x_1 x_2 + x_2^2)$ and relative entropy $e_1(u_0|M_\alpha) = -\ln(2(x_2 - x_1)) + \frac{1}{2}\ln(2\pi\alpha) + \frac{1}{2}$. Choosing $x_1 = 1$ and $x_2 = 1.1$, we get $\alpha \approx 1.1033$, $e_1(u_0|M_\alpha) \approx 3.0775$ and $E_{\min}^{\infty} \approx 0.0023$. The limit (2.27) of the ratio of the improved decay estimate to the original one is approximately 0.1621. Fig. 3 shows a comparison of the two decay estimates: the original (2.13) and the improved (2.23).

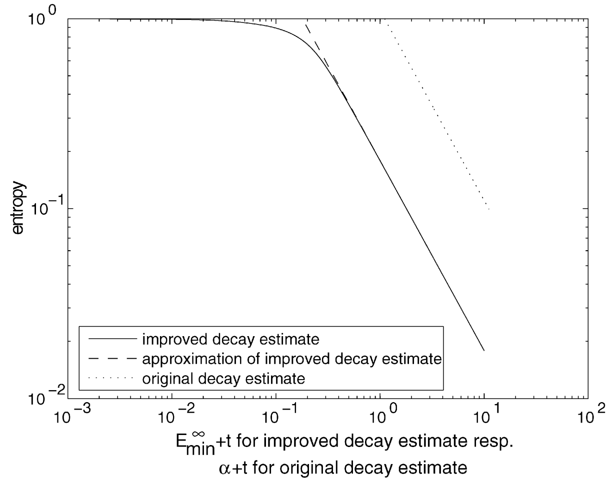
Remark 6 *(L^1-decay).* Using the well-known Csiszár–Kullback inequality for probability densities [12,21]

$$\|\rho_1 - \rho_2\|_{L^1(\mathbb{R}^n)}^2 \leqslant 2e_1(\rho_1|\rho_2),$$

the above decay estimates imply analogous results in L^1.

Remark 7 *(More general parabolic equations).* In the spirit of [26], we can extend Theorem 4 to derive improved decay estimates for uniformly parabolic equations of the form

$$\frac{\partial u}{\partial t} = \frac{k}{2}\operatorname{div}\big(\big[\mathbb{I} + A(x,t)\big]\nabla u\big).$$



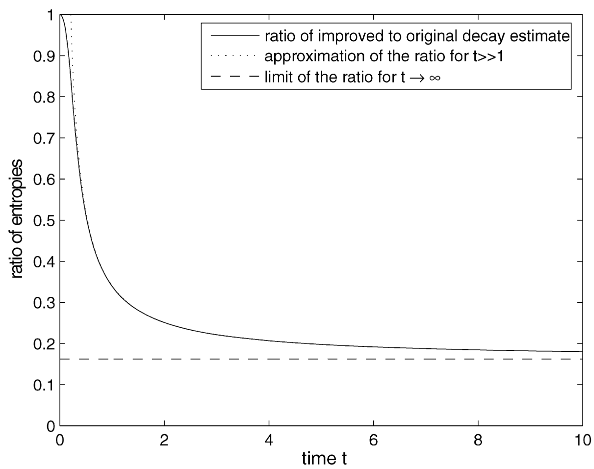


Fig. 3. Left: Logarithmic plot of the original decay estimate (2.13), the improved decay estimate (2.23), and the approximation (2.25) of the improved decay estimate for $t \gg 1$ divided by $e_1(u_0|M_\alpha)$. The original decay estimate is plotted against $\alpha + t$ and the improved decay estimate as well as its approximation against $E^{\infty}_{\min} + t$ in order to verify its asymptotic behaviors $(\alpha + t)^{-1}$ and $(E^{\infty}_{\min} + t)^{-1}$, respectively. Right: Ratio of the original (2.13) to the improved decay estimate (2.23), the approximation (2.26) of the ratio for $t \gg 1$, and the limit (2.27) of this approximation for $t \to \infty$.

Here, $\mathbb{I}$ is the identity matrix on $\mathbb{R}^n$. The symmetric positive perturbation A as well as its first spatial derivative are supposed to decay like $(1+t)^{-\beta}$ for some $0 < \beta < 1$ (cf. [20,26] for details).

Remark 8 *(Fokker–Planck equations).* Using the time dependent rescaling

$$u(\xi, \tau) = R(\tau)^{-n} v\left(\frac{\xi}{R(\tau)}, \ln R(\tau)\right)$$

with $R(\tau) = \sqrt{k\tau + 1}$ transforms the heat equation (2.1) into the *Fokker–Planck equation*

$$\frac{\partial v}{\partial t} = \operatorname{div}(\nabla v + xv), \quad x \in \mathbb{R}^n,\ t > 0.$$

Hence, the new decay estimate from Theorem 4 immediately translates into an improved decay estimate for the solution of the Fokker–Planck equation toward its unique normalized steady state $v_\infty(x) = e^{-|x|^2/2}$ (cf. [20] for details).

3. The heat equation and non-isotropic Gaussians

Up to now we have considered decay estimates for the heat equation w.r.t. isotropic Gaussians of the shape (2.2). Various convergence rates for the heat equation to more universal Gaussian densities in $\mathbb{R}^n$, namely Gaussians with an arbitrary covariance matrix, were found in [16]. Motivated by these results we shall now generalize our convergence rates in logarithmic entropy of Section 2 to solutions for the heat equation with respect to general non-isotropic Gaussians:

$$M_\Sigma(x) := (2\pi)^{-n/2}(\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2}\Sigma^{-1}x \cdot x\right), \quad x \in \mathbb{R}^n, \tag{3.1}$$

where the covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ is symmetric and positive definite.

Given an initial probability density u_0 on $\mathbb{R}^n$ with finite second moments and an arbitrary $x_0 \in \mathbb{R}^n$, we define the positive definite matrix $K(t) = (K_{ij}(t))_{i,j=1,\dots,n}$ for all times $t \geqslant 0$ by

$$K_{ij}(t) := \int_{\mathbb{R}^n} (x - x_0)_i (x - x_0)_j u(x,t)\, dx, \quad i, j = 1, \dots, n, \tag{3.2}$$

where u is the solution of the heat equation (2.1). It is a simple matter to check that the evolution of the second moments of the solution is linear in time, more precisely, $K(t) = K(0) + kt\mathbb{I}$ for all $t \geqslant 0$.

Let us remark that the n-dimensional Gaussian density with first moment x_0 and covariance matrix $K(t)$, $M_{K(t)}(x - x_0)$ is a solution itself to the heat equation (2.1). Since the heat equation is invariant under rotation of the coordinate system, we shall assume w.r.o.g. that $K(0)$ is diagonal. Since $K(t) = K(0) + kt\mathbb{I}$ is then diagonal for all $t \geqslant 0$. Hence, $M_{K(t)}(x - x_0)$ is a tensor product of 1D Gaussians, each of which satisfies the 1D heat equation.

3.1. Decay in relative entropy

The linear growth in time of the covariance matrix of general solutions motivates to consider its entropy behavior w.r.t. Gaussians with a covariance matrix of the form $\mathbb{E} + kt\mathbb{I}$, where $\mathbb{E} \in \mathbb{R}^{n\times n}$ is an arbitrary positive definite and symmetric matrix. We start with a lemma that is similar to [26, Lemma 1]:

Lemma 9 *(Finite relative Fisher information). Let the initial value $u_0 \in C(\mathbb{R}^n) \cap W^{1,2}_{\mathrm{loc}}(\mathbb{R}^n) \cap L^1_+(\mathbb{R}^n)$ be a probability density on $\mathbb{R}^n$ with finite second moments. Then there is a constant $C > 0$ such that the solution u of the IVP* (2.1) *satisfies*

$$I\big(u(t)\big|u_\infty(t)\big) < C, \quad \forall t \in [t_1, t_2],\ 0 < t_1 < t_2 < \infty, \tag{3.3}$$

with $u_\infty(x,t) := M_{\mathbb{E}+kt\mathbb{I}}(x - x_0)$ for an arbitrary $x_0 \in \mathbb{R}^n$ and an arbitrary symmetric and positive definite matrix $\mathbb{E} \in \mathbb{R}^{n\times n}$. Moreover,

$$\lim_{|x_j|\to\infty} \frac{\partial u(t)}{\partial x_j}\left(\ln\frac{u(t)}{u_\infty(t)} + 1\right) = 0, \tag{3.4}$$

$$\lim_{|x_j|\to\infty} \frac{\partial u_\infty(t)}{\partial x_j}\frac{u(t)}{u_\infty(t)} = 0, \tag{3.5}$$

for all $t > 0$ and $j = 1,\dots,n$.

Proof. We calculate for all $t > 0$,

$$\frac{\partial}{\partial x_i}\frac{u}{u_\infty} = \frac{1}{u_\infty}\frac{\partial u}{\partial x_i} + \frac{u}{u_\infty}\big((\mathbb{E} + kt\mathbb{I})^{-1}x\big)_i$$

and find

$$\left|\nabla\frac{u}{u_\infty}\right|^2 = \frac{1}{(u_\infty)^2}|\nabla u|^2 + \frac{u^2}{(u_\infty)^2}\big|(\mathbb{E} + kt\mathbb{I})^{-1}x\big|^2 + \frac{u}{(u_\infty)^2}2\big((\mathbb{E} + kt\mathbb{I})^{-1}x\big)\cdot\nabla u.$$

This leads to

$$I\big(u(t)\big|u_\infty(t)\big) = \int_{\mathbb{R}^n}\frac{1}{u}|\nabla u|^2\,dx + \int_{\mathbb{R}^n}\big|(\mathbb{E} + kt\mathbb{I})^{-1}x\big|^2 u\,dx + 2\int_{\mathbb{R}^n}(\mathbb{E} + kt\mathbb{I})^{-1}x\cdot\nabla u\,dx, \tag{3.6}$$

and the last term equals $-2\,\mathrm{tr}(\mathbb{E} + kt\mathbb{I})^{-1}$.

It is proved in [26] that the first integral in the expression (3.6) is bounded for an initial function $u_0 \in C(\mathbb{R}^n) \cap W^{1,2}_{\mathrm{loc}}(\mathbb{R}^n)$,

$$\int_{\mathbb{R}^n}\frac{1}{u}|\nabla u|^2\,dx \leqslant \int_{\mathbb{R}^n}\frac{1}{M_{kt}}|\nabla M_{kt}|^2\,dx = \frac{n}{kt} \leqslant \frac{n}{kt_1}, \quad t \geqslant t_1.$$

Hence,

$$I\big(u(t)\big|u_\infty(t)\big) = \int_{\mathbb{R}^n}\frac{1}{u}|\nabla u|^2\,dx + \int_{\mathbb{R}^n}\big|(\mathbb{E} + kt\mathbb{I})^{-1}x\big|^2 u\,dx - 2\,\mathrm{tr}(\mathbb{E} + kt\mathbb{I})^{-1} < C, \tag{3.7}$$

for all $t \geqslant t_1 > 0$. In [26] Toscani also showed

$$\lim_{|x_j|\to\infty} \frac{\partial u}{\partial x_j}(\ln u + 1) = 0.$$

Since u and $\frac{\partial u}{\partial x_j}$ are smooth, fast-decaying at infinity functions and have finite second-order moments, we deduce

$$-\lim_{|x_j|\to\infty} \frac{\partial u}{\partial x_j} \ln u_\infty = \frac{1}{2} \lim_{|x_j|\to\infty} \frac{\partial u}{\partial x_j}\big[(\mathbb{E}+kt\mathbb{I})^{-1}x\cdot x + C(t)\big] = 0$$

and

$$\lim_{|x_j|\to\infty} \frac{\partial u_\infty}{\partial x_j}\frac{u}{u_\infty} = -\lim_{|x_j|\to\infty} \big((\mathbb{E}+kt\mathbb{I})^{-1}x\big)_j u = 0,$$

concluding the proof. □

This lemma leads to the proof of a decay rate in relative entropy for the solution of the heat equation w.r.t. general Gaussians.

Theorem 10 *(Basic decay estimate). Let the initial value for the heat equation $u_0 \in C(\mathbb{R}^n)\cap W^{1,2}_{\mathrm{loc}}(\mathbb{R}^n)\cap L^1_+(\mathbb{R}^n)$ be a probability density on $\mathbb{R}^n$ with finite second moment and entropy. Then the relative entropy of the solution u to the IVP* (2.1) *w.r.t. $u_\infty(x,t) := M_{\mathbb{E}+kt\mathbb{I}}(x-x_0)$ with an arbitrary $x_0\in\mathbb{R}^n$ and an arbitrary symmetric and positive definite matrix $\mathbb{E}\in\mathbb{R}^{n\times n}$ converges to zero as $t\to\infty$. More precisely,*

$$e_1\big(u(t)\big|u_\infty(t)\big) \leqslant \frac{\rho(\mathbb{E})}{\rho(\mathbb{E})+nkt} e_1\big(u_0|u_\infty(0)\big), \quad \forall t\geqslant 0, \tag{3.8}$$

where $\rho(\mathbb{E})$ denotes the spectral radius of $\mathbb{E}$.

Proof. Following the proof of Theorem 1 we obtain for all $t>0$,

$$\begin{aligned}
\frac{d}{dt}e_1\big(u(t)\big|u_\infty(t)\big) &= \int_{\mathbb{R}^n} \frac{\partial u}{\partial t}\left[\ln\left(\frac{u}{u_\infty}\right)+1\right]dx - \int_{\mathbb{R}^n} \frac{\partial u_\infty}{\partial t}\frac{u}{u_\infty}\,dx \\
&= \frac{k}{2}\int_{\mathbb{R}^n} \Delta u\left[\ln\left(\frac{u}{u_\infty}\right)+1\right]dx - \frac{k}{2}\int_{\mathbb{R}^n} \Delta u_\infty \frac{u}{u_\infty}\,dx \\
&= -\frac{k}{2}\int_{\mathbb{R}^n} \frac{u_\infty}{u}\nabla u\cdot\nabla\left(\frac{u}{u_\infty}\right)dx + \frac{k}{2}\int_{\mathbb{R}^n} \nabla u_\infty\cdot\nabla\frac{u}{u_\infty}\,dx \\
&= -\frac{k}{2} I\big(u(t)\big|u_\infty(t)\big),
\end{aligned} \tag{3.9}$$

where $I(u(t)|u_\infty(t))$ is the relative Fisher information (2.5) of u w.r.t. u_∞. We conclude that the relative entropy $e_1(u(t)|u_\infty(t))$ is monotonically decreasing w.r.t. time. In the above integrations by parts, the boundary terms disappear due to Lemma 9.

Next we shall apply a LSI for the measure $u_\infty(t)$. To this end we use $\rho(\mathbb{E})\mathbb{I}\geqslant\mathbb{E}>0$ and hence

$$(\mathbb{E}+kt\mathbb{I})^{-1} \geqslant \big(\rho(\mathbb{E})+kt\big)^{-1}\mathbb{I}, \quad t\geqslant 0. \tag{3.10}$$

Since

$$\mathrm{Hess}_x\big[-\ln u_\infty(x,t)\big] = \mathrm{Hess}_x\big[-\ln M_{\mathbb{E}+kt\mathbb{I}}(x-x_0)\big] = (\mathbb{E}+kt\mathbb{I})^{-1},$$

(3.10) shows that $u_\infty(t)$ is uniformly log-concave with lower bound $(\rho(\mathbb{E})+kt)^{-1}$. Thus, (3.10) is a *Bakry–Emery condition* for the probability density $u_\infty(t)$, cf. [3–5]. Hence, $u_\infty(t)$ satisfies the LSI

$$e_1\big(\rho\big|u_\infty(t)\big) \leqslant \frac{\rho(\mathbb{E})+kt}{2} I\big(\rho\big|u_\infty(t)\big) \tag{3.11}$$

$\forall\rho\in L^1_+(\mathbb{R}^n)$ with $\int\rho\,dx=1$. Combining (3.9) and (3.11) yields

$$\frac{d}{dt}e_1\big(u(t)\big|u_\infty(t)\big) \leqslant -\frac{k}{\rho(\mathbb{E})+kt}e_1\big(u(t)\big|u_\infty(t)\big), \quad t>0,$$

and Gronwall's lemma implies the decay estimate (3.8). □

3.2. Improved decay estimate in relative entropy

Following the strategy of Section 2 we shall improve the decay estimate of Theorem 10 for the solution of the heat equation. In a first step we identify, for each fixed $t \geqslant 0$, the optimal non-isotropic Gaussian in the same sense as in the case of standard Gaussians. This shall yield by a minimization method an improvement of the convergence rate in relative entropy w.r.t. general Gaussians.

We consider an initial probability density u_0 on $\mathbb{R}^n$ with its center of mass at $x_0 \in \mathbb{R}^n$, i.e. $\int (x-x_0)u_0\,dx = 0$, and with $\int |x-x_0|^2 u_0\,dx < \infty$. For all $t \geqslant 0$ the covariance matrix $K(t) \in \mathbb{R}^{n\times n}$ of the solution $u(t)$ to the heat equation is defined by

$$K_{ij}(t) := \int_{\mathbb{R}^n} (x-x_0)_i (x-x_0)_j u(x,t)\,dx, \quad i,j=1,\dots,n. \tag{3.12}$$

As in previous section, we know that $K(t) = K(0) + kt\mathbb{I}$.

Now we want to find the general Gaussian $u_\infty(x,t) := M_{\bar{\mathbb{E}}(t)}(x-\bar{x}_0)$ that minimizes (for each fixed $t \geqslant 0$) the relative entropy $e_1(u(t)|u_\infty(t))$. The optimal first moment $\bar{x}_0 \in \mathbb{R}^n$ of u_∞ and the optimal positive definite matrix $\bar{\mathbb{E}}(t) \in \mathbb{R}^{n\times n}$ are such that the 0th and 1st moments of $u(t)$ and $u_\infty(t)$, as well as their covariance matrices coincide:

Theorem 11 *(Optimal non-isotropic Gaussian). Let the initial value u_0 be a probability density on $\mathbb{R}^n$ with finite second moment and entropy. Then, for each fixed time $t \geqslant 0$, $e_1(u(t)|M_{K(0)+kt\mathbb{I}}(\cdot - x_0))$ is the smallest relative entropy of the solution u of the IVP* (2.1) *w.r.t. all general Gaussians $M_{\tilde{\mathbb{E}}(t)}(x-\tilde{x}_0)$ with an arbitrary $\tilde{x}_0 \in \mathbb{R}^n$ and an arbitrary positive definite matrix $\tilde{\mathbb{E}}(t) \in \mathbb{R}^{n\times n}$.*

Proof. The relative entropy reads

$$e_1\big(u(t)\big|M_{\tilde{\mathbb{E}}(t)}(\cdot-\tilde{x}_0)\big) = \int_{\mathbb{R}^n} u\ln u\,dx + \frac{n}{2}\ln\big(2\pi\big[\det\tilde{\mathbb{E}}(t)\big]^{1/n}\big) + \frac{1}{2}\int_{\mathbb{R}^n} \tilde{\mathbb{E}}(t)^{-1}(x-\tilde{x}_0)\cdot(x-\tilde{x}_0)u\,dx. \tag{3.13}$$

$f(\tilde{x}_0)$ defined as the third term on the right-hand side of this equation is minimal w.r.t. $\tilde{x}_0 \in \mathbb{R}^n$ if and only if

$$\nabla f(\tilde{x}_0) = -\int_{\mathbb{R}^n} \tilde{\mathbb{E}}(t)^{-1}(x-\tilde{x}_0)u(x,t)\,dx = 0.$$

Since the matrix $\tilde{\mathbb{E}}(t)$ is regular, this condition is equivalent to

$$\int_{\mathbb{R}^n} (x-\tilde{x}_0)u(x,t)\,dx = 0. \tag{3.14}$$

Since $u(x,t)$ conserves the center of mass (cf. (2.10)) we conclude: The relative entropy $e_1(u(t)|M_{\tilde{\mathbb{E}}(t)}(\cdot-\tilde{x}_0))$ is minimal w.r.t. $\tilde{x}_0 \in \mathbb{R}^n$ iff the first moments of u and $M_{\tilde{\mathbb{E}}(t)}(x-\tilde{x}_0)$ coincide, i.e. iff $x_0 = \tilde{x}_0$.

To determine the positive definite matrix $\bar{\mathbb{E}} \in \mathbb{R}^{n\times n}$ minimizing, for each fixed $t \geqslant 0$, the relative entropy (3.13), we have to minimize

$$\ln(\det\tilde{\mathbb{E}}) + \mathrm{tr}\big(\tilde{\mathbb{E}}^{-1}K\big)$$

w.r.t. all positive definite matrices $\tilde{\mathbb{E}} = \tilde{\mathbb{E}}(t)$. $K = K(0) + kt\mathbb{I}$ denotes here the covariance matrix of $u(t)$. To simplify the computation we put $\mathbb{F} := \sqrt{K}\tilde{\mathbb{E}}^{-1}\sqrt{K} \geqslant 0$. Using the cyclicity of the trace we now have to minimize

$$-\ln(\det\mathbb{F}) + \mathrm{tr}(\mathbb{F}) = \sum_{j=1}^n (\lambda_j - \ln\lambda_j)$$

w.r.t. all positive definite matrices $\mathbb{F}$, with λ_j denoting its eigenvalues. Clearly, the unique minimum is attained at $\mathbb{F}=\mathbb{I}$, or equivalently, at $\bar{\mathbb{E}}=K$. $\square$

With this knowledge of the optimal Gaussian we are able to improve our decay estimate (3.8) for the solution of the heat equation w.r.t. general Gaussians, using a method similar to the one of Section 2. Theorem 10 yields for the solution $u(t)$ of the heat equation (2.1) the decay estimate

$$e_1\big(u(t)\big|M_{\mathbb{E}+kt\mathbb{I}}(\cdot-\tilde{x}_0)\big)\leqslant\frac{\rho(\mathbb{E})}{\rho(\mathbb{E})+nkt}e_1\big(u_0\big|M_{\mathbb{E}}(\cdot-\tilde{x}_0)\big) \tag{3.15}$$

for an arbitrary $\tilde{x}_0\in\mathbb{R}^n$ and an arbitrary positive definite matrix $\mathbb{E}\in\mathbb{R}^{n\times n}$. We estimate like in (2.16),

$$\begin{aligned}e_1\big(u(t)\big|M_{K(0)+kt\mathbb{I}}(\cdot-x_0)\big)&=\min_{\substack{\mathbb{E}>0\\ \tilde{x}_0\in\mathbb{R}^n}}e_1\big(u(t)\big|M_{\mathbb{E}+kt\mathbb{I}}(\cdot-\tilde{x}_0)\big)\leqslant\inf_{\substack{\mathbb{E}>0\\ \tilde{x}_0\in\mathbb{R}^n}}\frac{\rho(\mathbb{E})}{\rho(\mathbb{E})+nkt}e_1\big(u_0\big|M_{\mathbb{E}}(\cdot-\tilde{x}_0)\big)\\ &=\inf_{\mathbb{E}>0}\frac{\rho(\mathbb{E})}{\rho(\mathbb{E})+nkt}e_1\big(u_0\big|M_{\mathbb{E}}(\cdot-x_0)\big).\end{aligned} \tag{3.16}$$

In the case $u_0(x)=M_{K(0)}(x-x_0)$ we find that $e_1(u(t)|M_{K(0)+kt\mathbb{I}}(\cdot-x_0))=0$ for all times $t\geqslant 0$ and it holds equality in (3.15). Hence, we obtain the minimum of the right-hand side of formula (3.16) for $\mathbb{E}=K(0)$, i.e. (3.15) is already optimal.

In the case $u_0(x)\neq M_{K(0)}(x-x_0)$ we have to minimize the function f, defined for each $t\geqslant 0$ on the cone of positive definite matrices

$$\begin{aligned}f(\mathbb{E},t)&:=\frac{\rho(\mathbb{E})}{\rho(\mathbb{E})+nkt}e_1\big(u_0|M_{\mathbb{E}}(\cdot-x_0)\big)\\ &=\frac{1}{2}\frac{\rho(\mathbb{E})}{\rho(\mathbb{E})+nkt}\bigg(2\int_{\mathbb{R}^n}u_0\ln u_0\,dx+n\ln(2\pi)+\ln(\det\mathbb{E})+\operatorname{tr}\big(\mathbb{E}^{-1}K(0)\big)\bigg)\geqslant 0\end{aligned} \tag{3.17}$$

w.r.t. $\mathbb{E}\geqslant 0$. This function has the following features.

Lemma 12 *(Computation of $\mathbb{E}_{\min}(t)$). For each fixed $t\geqslant 0$ the function $f(\mathbb{E},t)$ has w.r.t. to all symmetric and positive definite matrices $\mathbb{E}\in\mathbb{R}^{n\times n}$ a unique minimum at $\mathbb{E}_{\min}(t)\in\mathbb{R}^{n\times n}$ with the following properties*:

(a) $\mathbb{E}_{\min}=\mathbb{E}_{\min}(t)$ *satisfies*

$$\mathbb{E}_{\min}=\min\big(K(0),\rho_{\min}\mathbb{I}\big),$$

where $\rho_{\min}=\rho(\mathbb{E}_{\min})\leqslant\rho(K(0))$ is its spectral radius.

(b) $\mathbb{E}_{\min}(0)=K(0)$.

(c) $\mathbb{E}_{\min}(t)$ *is monotonically decreasing w.r.t. $t\geqslant 0$, i.e. $\mathbb{E}_{\min}(t_2)\leqslant\mathbb{E}_{\min}(t_1)$ for $0\leqslant t_1<t_2$ in the sense of positive definite matrices. In particular,*

$$0<\mathbb{E}_{\min}^{\infty}\leqslant\mathbb{E}_{\min}(t)\leqslant K(0),\quad\forall t\geqslant 0.$$

Proof. To simplify the notation we put $K=K(0)$. W.r.o.g. we shall assume $K=\operatorname{diag}(k_1,\dots,k_n)$ with $0<k_1\leqslant k_2\leqslant\cdots\leqslant k_n$. Indeed, if the minimum of $f_K(.,t)$ is attained at $\mathbb{E}=\mathbb{E}_{\min}$, the minimum of $f_{\tilde{K}}(.,t)$ with $\tilde{K}:=SKS^{-1}$ and S orthogonal is attained at $\tilde{\mathbb{E}}_{\min}:=S\mathbb{E}_{\min}S^{-1}$. This follows from $\rho(\mathbb{E}_{\min})=\rho(\tilde{\mathbb{E}}_{\min})$, $\det(\mathbb{E}_{\min})=\det(\tilde{\mathbb{E}}_{\min})$, $\operatorname{tr}(\mathbb{E}_{\min}^{-1}K)=\operatorname{tr}(\tilde{\mathbb{E}}_{\min}^{-1}\tilde{K})$, and hence $f_K(\mathbb{E}_{\min},t)=f_{\tilde{K}}(\tilde{\mathbb{E}}_{\min},t)$.

We shall now minimize $f(\mathbb{E},t)$ w.r.t. $\mathbb{E}$ positive definite and symmetric in three steps, to show that we can reduce to minimize a function of the spectrum of $\mathbb{E}$,

$$\begin{aligned}\min_{\mathbb{E}>0}f(\mathbb{E},t)&=\min_{\rho>0}f_1(\rho,t)\Big[\min_{\substack{\mathbb{E}>0\\ \rho(\mathbb{E})=\rho}}f_2(\mathbb{E})\Big]\\ &=\min_{\rho>0}f_1(\rho,t)\Big\{\min_{\substack{\sigma(\mathbb{E})\subset\mathbb{R}^+\\ \rho(\mathbb{E})=\rho\text{ fixed}}}\Big[\beta+\ln(\det\mathbb{E})+\min_{\substack{\mathbb{E}>0\\ \sigma(\mathbb{E})\subset\mathbb{R}^+\text{ fixed}}}\operatorname{tr}\big(\mathbb{E}^{-1}K\big)\Big]\Big\},\end{aligned} \tag{3.18}$$

with the scalar functions

$$f_1(\rho,t) := \frac{1}{2}\frac{\rho}{\rho+nkt}, \quad \rho>0,$$
$$f_2(\mathbb{E}) := \beta + \ln(\det\mathbb{E}) + \operatorname{tr}\big(\mathbb{E}^{-1}K\big) = \beta + \sum_{j=1}^{n}\left(\ln e_j + \frac{k_j}{e_j}\right), \quad \mathbb{E}>0, \tag{3.19}$$

with $\beta := 2\int u_0 \ln u_0\,dx + n\ln(2\pi)$ and $\sigma(\mathbb{E}) = \{0<e_1\leqslant\cdots\leqslant e_n=\rho\}$.

Step 1. First we shall minimize $\operatorname{tr}(\mathbb{E}^{-1}K)$ over all symmetric matrices $\mathbb{E}$ having the fixed spectrum $\sigma(\mathbb{E}) = \{0 < e_1\leqslant\cdots\leqslant e_n=\rho\}$. Since K is diagonal, the minimum of $\operatorname{tr}(\mathbb{E}^{-1}K)$ is attained at $\mathbb{E}_3 = \operatorname{diag}(e_1,\dots,e_n)$. Since the entries of K are increasing, also the e_j's have to increase. This is a direct consequence of the following result [25, Theorem 1]: *For all real symmetric matrices A, B it holds*

$$\sum_{j=1}^{n}\lambda_{n-j+1}(A)\lambda_j(B) \leqslant \operatorname{tr}(AB) \leqslant \sum_{j=1}^{n}\lambda_j(A)\lambda_j(B),$$

where the eigenvalues are labeled in increasing order. The right inequality is actually a special case of the *von Neumann trace inequality* [23]. The left inequality now yields the assertion

$$\operatorname{tr}\big(\mathbb{E}^{-1}K\big) \geqslant \sum_{j=1}^{n}\lambda_{n-j+1}\big(\mathbb{E}^{-1}\big)\lambda_j(K) = \sum_{j=1}^{n}\frac{k_j}{e_j} = \operatorname{tr}\big(\mathbb{E}_3^{-1}K\big) \quad \forall\mathbb{E}>0 \text{ with } \sigma(\mathbb{E}) \text{ fixed.}$$

Step 2. Next we minimize $f_2(\mathbb{E})$ over all diagonal matrices $\mathbb{E}>0$ subject to the constraint $\rho(\mathbb{E})=\rho$ with $\sigma(\mathbb{E}) = \{0<e_1\leqslant\cdots\leqslant e_n=\rho\}$. From this we can conclude that the unique minimum of $f_2(\mathbb{E})$ w.r.t. $\mathbb{E}>0$ and $\rho(\mathbb{E})=\rho$ (with ρ fixed) is attained at $\mathbb{E}_2 = (e_{21},\dots,e_{2n})$ with

$$e_{2j} = \min(k_j,\rho), \quad j\leqslant n-1,$$
$$e_{2n} = \rho. \tag{3.20}$$

We first remark that each summand of (3.19) is a decreasing function of e_j for $0<e_j<k_j$, increasing for $e_j>k_j$ achieving its minimum at $\bar e_j = k_j$. Now, the largest eigenvalue e_{2n} must be equal to ρ by definition of the minimization set of matrices. Taking into account both facts we verify (3.20). Using (3.20) in (3.19), the minimum of $f_2(\mathbb{E})$ satisfies

$$f_3(\rho) := \min_{\substack{\mathbb{E}>0\\ \rho(\mathbb{E})=\rho}} f_2(\mathbb{E}) = f_2(\mathbb{E}_2) = \beta + \ln\rho + \frac{k_n}{\rho} + \sum_{j=1}^{n-1}g_j(\rho), \quad \rho>0, \tag{3.21}$$

with the $C^1(\mathbb{R}^+)$-functions, $j=1,\dots,n-1$,

$$g_j(\rho) := \begin{cases} \ln\rho + \frac{k_j}{\rho}, & \rho\leqslant k_j,\\ \ln k_j + 1, & \rho\geqslant k_j. \end{cases} \tag{3.22}$$

Step 3. Next we minimize $f_1(\rho)f_3(\rho)$ w.r.t. $\rho>0$. This yields the following condition for $\rho_{\min}$:

$$nkt\big(f_3+\rho f_3'\big) = -\rho^2 f_3'. \tag{3.23}$$

Here,

$$f_3(\rho) + \rho f_3'(\rho) = \beta + \ln\rho + 1 + \sum_{j=1}^{n-1}\big(g_j+\rho g_j'\big),$$

with

$$g_j + \rho g_j' = \begin{cases} \ln\rho + 1, & \rho\leqslant k_j,\\ \ln k_j + 1, & \rho\geqslant k_j. \end{cases} \tag{3.24}$$

Hence, $f_3 + \rho f_3'$ is strictly monotonic increasing in $\rho > 0$,

$$\lim_{\rho\to 0+} f_3(\rho) + \rho f_3'(\rho) = -\infty$$

and

$$f_3(k_n) + k_n f_3'(k_n) = 2e\big(u_0 \big| M_{K(0)}(\cdot - x_0)\big) \geqslant 0.$$

On the other hand,

$$-\rho^2 f_3'(\rho) = k_n - \rho + \sum_{j=1}^{n-1} (k_j - \rho) H(k_j - \rho)$$

(with H denoting the Heaviside function) is strictly monotonic decreasing in $\rho > 0$, positive on $[0, k_n)$, and it has a unique zero at $\rho = k_n$. This implies that Eq. (3.23) has a unique solution $\rho_{\min}$ with $0 < \rho_{\min} \leqslant k_n$.

One easily checks that $\lim_{\rho\to 0+}(f_1 f_3)'(\rho) = -\infty$. Moreover, f_1 and f_3 are both strictly increasing on $[k_n, \infty)$. Hence, $(f_1 f_3)(\rho)$ takes its unique *minimum* at $\rho = \rho_{\min} \leqslant k_n$.

(a) From (3.20) we hence conclude $\mathbb{E}_{\min} = \min(K(0), \rho_{\min}\mathbb{I})$.

(b) $t = 0$ implies $\rho_{\min} = k_n$ and hence $\mathbb{E}_{\min}(0) = K(0)$.

(c) For $u_0 \neq M_{K(0)}(\cdot - x_0)$, the monotonicity properties of both sides of (3.23) imply that $\rho_{\min}(t)$ is strictly decreasing in t, with

$$0 < \rho_{\min}^{\infty} < \rho_{\min}(t) < k_n, \quad \forall t > 0.$$

Here, $\rho_{\min}^{\infty}$ is the unique minimum of $f_3 + \rho f_3'$ (cf. to the analogous situation in Lemma 3(d) and in Fig. 1). Hence, (3.20) implies that the matrix $\mathbb{E}_{\min}(t)$ is decreasing w.r.t. t. In the case $u_0(x) = M_{K(0)}(x - x_0)$ we have $\mathbb{E}_{\min}(t) = K(0)$. $\square$

Remark 13 *(Radial symmetric case).* In the special case of a radially symmetric initial condition with covariance matrix $K(0) = \frac{\alpha}{n} \cdot \mathbb{I}$, the above Lemma 12 reduces to Lemma 3 with $\mathbb{E}_{\min}(t) = E_{\min}(t) \cdot \mathbb{I}$ for $t \geqslant 0$. Then, the condition (3.23) is equivalent to

$$\bigg(\int_{\mathbb{R}^n} u_0(x) \ln u_0(x)\, dx + \frac{n}{2} \ln\big(2\pi E_{\min}(t)\big) + \frac{n}{2}\bigg) kt = \frac{\alpha}{2} - \frac{n}{2} E_{\min}(t).$$

Lemma 12 and Theorem 11 now directly yield an improved decay estimate in logarithmic entropy (compared to the result of Theorem 10):

Theorem 14 *(Improved decay estimate). Let $u_0 \in C(\mathbb{R}^n) \cap W^{1,2}_{\mathrm{loc}}(\mathbb{R}^n) \cap L^1_+(\mathbb{R}^n)$ be a probability density on $\mathbb{R}^n$ with finite second moment and entropy. Then the solution u of the IVP* (2.1) *satisfies*

$$e_1\big(u(t) \big| M_{K+kt\mathbb{I}}(\cdot - x_0)\big) \leqslant \frac{\rho(\mathbb{E}_{\min}(t))}{\rho(\mathbb{E}_{\min}(t)) + nkt} e_1\big(u_0 \big| M_{\mathbb{E}_{\min}(t)}(\cdot - x_0)\big) = f\big(\mathbb{E}_{\min}(t), t\big) \tag{3.25}$$

with $\mathbb{E}_{\min}(t)$ from Lemma 12 *and Remark* 13.

In particular, if $K(0) = \frac{\alpha}{n} \cdot \mathbb{I}$ we have $\mathbb{E}_{\min}(t) = E_{\min}(t) \cdot \mathbb{I}$ for $t \geqslant 0$, and Theorem 14 reduces to Theorem 4 for standard Gaussians.

Acknowledgments

A.A. was partially supported by the DFG under grant No. AR 277/3-3 and by the ESF in the project "Global and geometrical aspects of nonlinear partial differential equations." J.A.C. acknowledges the support from DGI-MEC (Spain) FEDER-project MTM2005-08024 and 2005SGR00611. A.A. and J.A.C. acknowledge partial support of the Acc. Integ. program HU2006-0025.

References

[1] S. Angenent, Large time asymptotics for the porous media equation, in: Nonlinear Diffusion Equations and Their Equilibrium States I, Berkeley, CA, 1986, in: Math. Sci. Res. Inst. Publ., vol. 12, Springer, 1988, pp. 21–34.

[2] A. Arnold, J.A. Carrillo, L. Desvillettes, J. Dolbeault, A. Jüngel, C. Lederman, P.A. Markowich, G. Toscani, C. Villani, Entropies and equilibria of many-particle systems: An essay on recent research, Monatsh. Math. 142 (2004) 35–43.

[3] A. Arnold, P. Markowich, G. Toscani, A. Unterreiter, On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Plack type equations, Comm. Partial Differential Equations 26 (2001) 43–100.

[4] D. Bakry, M. Emery, Hypercontractivité de semi-groupes de diffusion, C. R. Acad. Sci. Paris Sér. I 15 (1984) 299–305.

[5] D. Bakry, M. Emery, Diffusions hypercontractives, in: Sém. Prob. XIX, in: Lecture Notes in Math., vol. 1123, Springer, 1985, pp. 177–206.

[6] G.I. Barenblatt, Scaling, Self-Similarity, and Intermediate Asymptotics, Cambridge Univ. Press, 1996, reprinted 1997.

[7] E.A. Carlen, Superadditivity of Fisher's information and logarithmic Sobolev inequalities, J. Funct. Anal. 101 (1991) 194–211.

[8] J.A. Carrillo, M. DiFrancesco, G. Toscani, Strict contractivity of the 2-Wasserstein distance for the porous medium equation by mass-centering, Proc. Amer. Math. Soc. 135 (2007) 353–363.

[9] J.A. Carrillo, A. Jüngel, P. Markowich, G. Toscani, A. Unterreiter, Entropy production methods for degenerate parabolic problems and generalized Sobolev inequalities, Monatsh. Math. 133 (2001) 1–82.

[10] J.A. Carrillo, G. Toscani, Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity, Indiana Univ. Math. J. 49 (2000) 113–142.

[11] C. Cercignani, H-theorem and trend to equilibrium in the kinetic theory of gases, Arch. Mech. 34 (1982) 231–241.

[12] I. Csiszár, Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis von Markoffschen Ketten, Magyar Tud. Akad. Mat. Kutato Int. Kl. 8 (1963) 85–108.

[13] J. Denzler, R.J. McCann, Fast diffusion to self-similarity: Complete spectrum, long-time asymptotics and numerology, Arch. Ration. Mech. Anal. 175 (2004) 301–342.

[14] J. Duoandikoetxea, E. Zuazua, Moments, masses de Dirac et décomposition de fonctions, C. R. Acad. Sci. Paris Sér. I Math. 315 (1992) 693–698.

[15] R. Fisher, Theory of statistical estimation, Math. Proc. Cambridge Philos. Soc. 22 (1925) 700–725.

[16] T. Goudon, S. Junca, G. Toscani, Fourier-based metrics and Berry–Esseen like inequalities, Monatsh. Math. 135 (2002) 115–136.

[17] L. Gross, Logarithmic Sobolev inequalities, Amer. J. Math. 97 (1975) 1061–1083.

[18] Y.-J. Kim, R.J. McCann, Sharp decay rates for the fastest conservative diffusions, C. R. Acad. Sci. Paris Sér. I Math. 341 (2005) 157–162.

[19] Y.-J. Kim, R.J. McCann, Potential theory and optimal convergence rates in fast nonlinear diffusion, J. Math. Pures Appl. 86 (2006) 42–67.

[20] C. Klapproth, Entropy methods for the large-time behavior of parabolic equations, Master thesis, Münster University, 2006.

[21] S. Kullback, Information Theory and Statistics, John Wiley, 1959.

[22] R.J. McCann, D. Slepcev, Second-order asymptotics for the fast-diffusion equation, Int. Math. Res. Not. 24947 (2006) 1–22.

[23] L. Mirsky, A trace inequality of John von Neumann, Monatsh. Math. 79 (1975) 303–306.

[24] F. Otto, The geometry of dissipative evolution equations: The porous medium equation, Comm. Partial Differential Equations 26 (2001) 101–174.

[25] H. Richter, Zur Abschätzung von Matrizennormen, Math. Nachr. 18 (1958) 178–187.

[26] G. Toscani, Kinetic approach to the asymptotic behaviour of the solution to diffusion equations, Rend. Mat. 16 (1996) 329–346.

[27] G. Toscani, Sur l'inégalité logarithmique de Sobolev, C. R. Acad. Sci. Paris Math. 324 (1997) 689–694.

[28] A. Unterreiter, A. Arnold, P. Markowich, G. Toscani, On generalized Csiszár–Kullback inequalities, Monatsh. Math. 131 (2000) 235–253.

[29] J.L. Vázquez, Asymptotic behaviour for the porous medium equation posed in the whole space, J. Evol. Equ. 3 (2003) 67–118.

[30] J.L. Vázquez, The Porous Medium Equation. Mathematical Theory, Oxford Univ. Press, 2007.

[31] T.P. Witelski, A.J. Bernoff, Self-similar asymptotics for linear and nonlinear diffusion equations, Stud. Appl. Math. 100 (1998) 153–193.